基于公有云的新闻服务平台架构设计

杨 旗
（新华通讯社，北京 100803）

摘 要：随着国际形势的风起云涌，如何加强我国在海外的传媒影响能力将显得尤其重要。本文基于公有云对新华社供稿服务架构进行了优化设计，全面提速稿件海外落地速度和新闻传播效果，提高海外用户访问体验，增强新华社的国际传播能力。

关键词：新闻服务；公有云；平台架构设计；用户访问体验　　**中图分类号：**TP237　　**文献标识码：**A
文章编号：1671-0134（2021）03-109-03　　**DOI：**10.19483/j.cnki.11-4653/n.2021.03.030
本文著录格式：杨旗 . 基于公有云的新闻服务平台架构设计 [J]. 中国传媒科技，2021（03）：109-110，113.

导语

目前国际形势风起云涌，加强我国在海外的传媒影响能力势在必行。而随着云计算技术和互联网的快速发展，传统媒体面临从“纸媒”到“屏媒”的用户之变。如今时代谁掌握了互联网，谁就把握住了时代主动权，而云计算技术又是互联网发展的重要基石，所以基于公有云的新闻服务平台架构应运而生。

1. 公有云发展现状

公有云服务是云服务厂商提供的一种面向公众的云资源共享服务。公众可按需购买公有云服务，其按服务模式可分为IaaS（基础设施即服务）、PaaS（平台即服务）、SaaS（软件即服务）三个服务层次。

（1）IaaS 层是将计算、网络、存储等资源虚拟化后以服务方式提供给公众。其是云计算中最基础的资源，也是云服务商最主要的收入来源。

（2）PssS 层是建立在 IaaS 层之上的平台层，即将中间件、数据库等资源封装后为应用程序提供良好的开发运行时环境。

（3）SaaS 层是公有云服务商将其成熟的应用软件部署于其云平台上提供服务。公众可通过其提供的软件接口或者通过 web 浏览器访问。用户购买 SaaS 层服务将极大地节约软件开发及运维成本，然而 SaaS 层服务是面向多租户的软件即服务，其在针对单一用户的个性化定制方面仍然存在不足。

目前全球云计算市场迅猛发展，尤其是 2020 年新冠肺炎疫情的爆发极大地刺激了公有云服务市场的投资。据 IDC 调查研究，很多服务商的云计算策略因受一季度疫情影响，都发生了变化，大概有 28% 的企业计划全部向公有云服务转移，48% 的企业计划将云服务部分向公有云服务转移。在第二季度，服务商在基础设施投资方面也出现明显变化，基于公有云服务的基础设施投资增长了 49 %，而基于传统 IT 的基础设施投资下降了 9%，IDC 根据这个调研数据预计 2022 年基于公有云方面的基础设施的建设投资将占据主导地位，超过传统的 IT 基础设施投资。[1]

2. 新闻服务平台架构设计

2.1　新闻服务平台

新闻服务平台（即供稿平台）是基于互联网的、面向新闻机构用户，提供包括所有新华社文字、图片、图表、视频、新媒体、多媒体、历史资料和第三方产品的在线供稿服务和信息数据在线推送服务的技术系统。

新闻服务平台主要的业务场景包含:（1）在线推送（2）在线浏览（3）用户订购信息管理（4）在线点题（5）产品在线打包和发布（6）用户管理（7）在线交流与通知（8）新媒体内容展示（9）在线检索（10）运营分析（11）第三方平台（12）服务能力监控。

新闻服务平台每日产生稿件数据总量大约为 40G。平台将存储 3 个月的生产数据。稿件数据分为稿件原始数据、稿件元数据。稿件原始数据为 XML 文件和附件文件。其中，XML 文件为稿件描述，稿件附件文件主要类型为图片、视频、音频、HTML/CSS/JS、PDF、CDR、ZIP 文件等。文字主要存储在 NoSQL 数据库和 MySQL 关系数据库，图片的实体存储在 NAS 存储，图片的元数据信息存储在 MySQL 关系数据库。其它附件元数据也同样存储在 MySQL 数据库，实体文件存储在 NAS 存储。图片除对外供稿外，需要进行中图和小图的抽图，用于网页展示。视频除对外供稿外，可能需要进行视频转码和关键帧抽取等。

新闻服务平台将通过客户端同步、EMail、RSS、API 接口、FTP 推送等方式，为海外机构用户提供支持 XinhuaML、CNML、NewsML、SimpleML、RSS、IPTC 、纯文本等格式的数据服务。

2.2　架构设计

新闻服务平台（即供稿平台）采用基于公有云的架构设计，用极短的时间搭建起以亚太（香港）、北美（美国）、欧洲（法国）三个站点为核心（如图 1 所示）、可扩展、覆盖全球的云供稿服务，实现稿件全球快速分发，为海外用户提供一站式云供稿服务。新闻服务平台

北京站点为新华社自有站点，其与亚太站点、北美站点、欧洲站点存储的稿件数据基本一致，缓存的热点稿件数据会因各地机构用户访问情况的不同而不同。用户在线下载、浏览、评论等行为数据将回传北京站点。

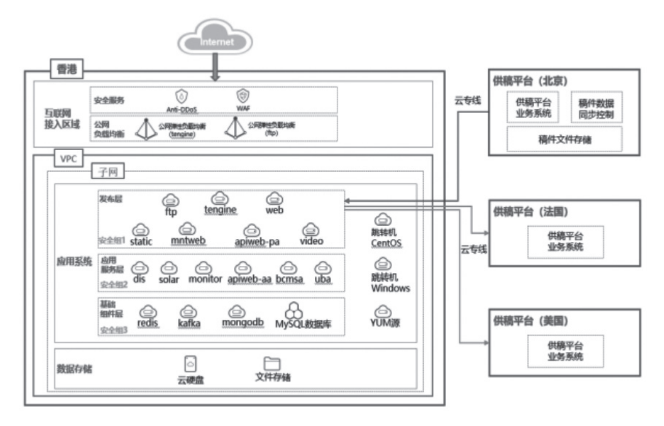


图 1 新闻服务平台架构

2.2.1 网络自定义

采用公有云的网络自定义能力，根据部署需要，自行定义、规划、管理网络，包括自定义划分子网地址、自定义安全组的 ACL 规则等。公有云中的虚拟私有云（VPC）可以在云端构建一个或多个隔离的网络环境，在保证网络安全的同时又满足了自主定义网络的诉求。

根据新闻服务平台需求特点，在亚太、北美、欧洲分别建立了一个单独的 VPC，同时将应用系统分为发布层、应用服务层、基础组件层，通过安全组技术，形成三个对应的安全组，保障业务的逻辑隔离，确保实现最小权限访问，保障安全。而互联网接入区采用公有云的防 DDoS 服务和 WAF 服务，互联网出口端则使用动态 BGP，保证通信的高可用性、低时延，不受运营商影响。

2.2.2 稿件数据同步

稿件数据的传输同步离不开基础传输网络环境，而国际互联网环境复杂不稳定，因此为保障稿件数据同步的稳定，新闻服务平台采用公有云专线接口服务将北京站点与亚太站点（即香港站点）通过两条冗余的专线连通，而亚太站点、欧洲站点、北美站点之间则通过公有云中的云专线服务互连。

稿件数据同步如图 2 所示，流程如下：

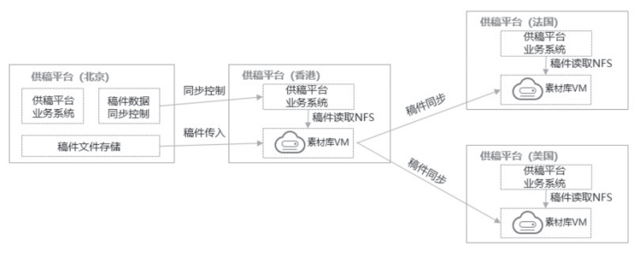


图 2 稿件数据同步

（1）在香港站点创建 MongoDB 数据库，用于存放稿件文件；

（2）北京主站的供稿系统，将主站的稿件数据同步上传到香港的数据库中；

（3）我国香港、法国和美国的供稿平台，都通过访问我国香港的数据库，来加载和获取稿件。

2.2.3 高可靠性设计

新闻服务平台的高可靠性设计首先是云平台自身的高可靠性设计：（1）配置反亲和性策略：反亲和性主要是出于高可靠性考虑，尽量分散实例，某个节点故障的时候，对应用的影响只是一个实例。云服务器组内的弹性云服务器将遵循反亲和策略，尽量分散地创建在不同主机上，提高可靠性保证。（2）云服务器 HA 自动恢复策略：当计算节点宕机或者虚拟机状态为 error，系统可以将具有 HA 属性的虚拟机故障迁移到其他健康的计算节点，保证虚拟机能够快速恢复。在创建云主机时，打开主机 HA 自动恢复功能，保障业务的高可靠性。（3）数据备份：采用虚拟机数据备份、管理数据备份等手段来保障高可靠。当需要恢复数据时，根据需要的备份文件，管理节点可以自动将数据恢复到备份时间点。其次是新闻服务平台系统多站点的高可靠性设计：通过全局负责均衡，海外用户不但能就近更高效地访问新闻服务平台，而且当某个站点出问题时，也能及时快速地切换至别的站点访问。

3. 应用效果及效益

公有云资源快速部署，相较于传统物理服务器部署方式，极大缩短部署时间，快速完成所有海外供稿用户的上云迁移。

基于公有云的新闻服务平台实现了全球站点统一运维，用户统一服务，提高供稿用户服务和响应能力。全球多站点集中维护，不仅大幅降低运维成本，而且提升了对海外营销的技术支撑保障力度。基于云节点全球覆盖面广的特性，新闻服务平台根据供稿系统的用户分布，灵活选择部署节点，使得用户能就近访问云供稿服务。

基于公有云的新闻服务平台具有灵活扩展优化的特性。根据云供稿服务实际的运行情况，对系统 CPU、内存、存储、网络资源进行灵活高效的扩展优化，保障云供稿服务时刻处在性能最佳状态，尤其是在重大保障期间，大量重要稿件的播出，用户访问激增，对基础环境的支撑要求尤为严格，公有云服务采用优先服务、重点保障的原则，提前优化扩充资源，为重保期间用户提供优质服务。

随着公有云服务不断优化，用户访问海外各站点较访问北京站点，web 系统和稿件 FTP 推送的平均延时缩短了 2–3 秒，速度提升了速度提高 60% 以上，取得了很好的社会效益，增强了我社供稿服务的形象和品牌竞争力。由于公有云服务实现了计算、存储、网

（下转第113页）